Brief Description of the Drawings

Figures 1A-1E show the nucleotide alignment of clones 4304443H1 (SEQ ID NO:1), 3040232H1 (SEQ ID NO:2) 3790941H1 (SEQ ID NO:3), 3424294H1 (SEQ ID NO:4) 2741038H1 (SEQ ID NO:5), 4302934H1 (SEQ ID NO:6), 158545H1 (SEQ ID NO:7), the full-length sequence of clone 4304443H1 [designated as 4304443inh (SEQ ID NO:8)], and the consensus sequence (SEQ ID NO:9) derived therefrom.

Figure 2 shows the contig map depicting the formation of the consensus nucleotide sequence (SEQ ID NO:9) from the nucleotide alignment of overlapping clones 4304443H1 (SEQ ID NO:1), 3040232H1 (SEQ ID NO:2) 3790941H1 (SEQ ID NO:3), 3424294H1 (SEQ ID NO:4) 2741038H1 (SEQ ID NO:5), 4302934H1 (SEQ ID NO:6), 158545H1 (SEQ ID NO:7), 4304443inh (SEQ ID NO:8).

Detailed Description of the Invention

The present invention provides a gene, or a fragment, thereof, which codes for a BS322 polypeptide having at least about 50% identity with SEQ ID NO:24 or SEQ ID NO:25. The present invention further encompasses a BS322 gene, or a fragment thereof, comprising DNA which has at least about 50% identity with SEQ ID NO:8 or SEQ ID NO:9.

The present invention also provides methods for assaying a test sample for products of a breast tissue gene designated as BS322, which comprises making cDNA from mRNA in the test sample, and detecting the cDNA as an indication of the presence of breast tissue gene BS322. The method may include an amplification step, wherein one or more portions of the mRNA from BS322 corresponding to the gene or fragments thereof, is amplified. Methods also are provided for assaying for the translation products of BS322. Test samples which may be assayed by the methods provided herein include tissues, cells, body fluids and secretions. The present invention also provides reagents such as oligonucleotide primers and polypeptides which are useful in performing these methods.

Portions of the nucleic acid sequences disclosed herein are useful as primers for the reverse transcription of RNA or for the amplification of cDNA; or as probes to

The following is the general procedure for the isolation and analysis of cDNA clones. In a particular embodiment disclosed herein, mRNA is isolated from breast tissue and used to generate the cDNA library. Breast tissue is obtained from patients by surgical resection and is classified as tumor or non-tumor tissue by a pathologist.

The cDNA inserts from random isolates of the breast tissue libraries are sequenced in party, analyzed in details as set forth in the Examples, and are disclosed in the Sequence Listing as SEQ ID NOS: 1-7. Also analyzed in detail as set forth in the Examples, and disclosed in the Sequence Listing is the full-length sequence of clone 4304443H1 [referred to herein as 4304443inh (SEQ ID NO:8)]. The consensus sequence of these inserts is presented as SEQ ID NO:9). These polynucleotides may contain an entire open reading frame with or without associated regulatory sequences for a particular gene, or them may encode only a portion of the gene of interest. This is attributed to the fact that many genes are several hundred and sometimes several thousand bases in length and, with current technology, cannot be cloned in their entirety because of vector limitations, incomplete reverse transcription of the first strand, or incomplete replication of the second strand. Contiguous, secondary clones containing additional nucleotide sequences may be obtained using a variety of methods known to those of skill in the art.

Methods for DNA sequencing are well known in the art. Conventional enzymatic methods employ DNA polymerase, Klenow fragment, Sequenase (US Biochemical Corp., Cleveland, OH), or Taq polymerase to extend DNA chains from an oligonucleotide primer annealed to the DNA template of interest. Methods have been developed for the use of both single-stranded and double-stranded templates. The chain termination reaction products may be electrophoresed on urea/polyacrylamide gels and detected either by autoradiography (for radionucleotide labeled precursors) or by fluorescence (for fluorescent-labeled precursors). Recent improvements in mechanized reaction preparation, sequencing and analysis using the fluorescent detection method have permitted expansion in the number of sequences that can be determined per day using machines such as the Applied Biosystems 377 DNA Sequencers (Applied Biosystems, Foster City, CA).

EXAMPLES

Example 1: Identification of Breast Tissue Library BS322 Gene-Specific Clones

A. Library Comparison of Expressed Sequence Tags (EST's) or Transcript Images. Partial sequences of cDNA clone inserts, so-called "expressed sequence tags" (EST's), were derived from cDNA libraries made from breast tumor tissues, breast non-tumor tissues and numerous other tissues, both tumor and non-tumor and entered into a database (LIFESEQ™ database, available from Incyte Pharmaceuticals, Palo Alto, CA) as gene transcript images. See International Publication No. WO95/20681. (A transcript image is a listing of the number of EST's for each of the represented genes in a given tissue library. EST's sharing regions of mutual sequence overlap are classified into clusters. A cluster is assigned a clone number from a representative 5' EST. Often, a cluster of interest can be extended by comparing its consensus sequence with sequences of other EST's which did not meet the criteria for automated clustering. The alignment of all available clusters and single EST's represent a contig from which a consensus sequence is derived.) The transcript images then were evaluated to identify EST sequences that were representative primarily of the breast tissue libraries. These target clones were then ranked according to their abundance (occurrence) in the target libraries and their absence from background libraries. Higher abundance clones with low background occurrence were given higher study priority. EST's corresponding to the consensus sequence of BS322 were found in 16.4% (9 of 55) of breast tissue libraries. EST's corresponding to the consensus sequence, SEQ ID NO:9 (or fragments thereof), were found in only 0.1% (1 of 940) of the other, non-breast, libraries of the database. Therefore, the consensus sequence, or fragments thereof, were found more than 148 times more often in breast than non-breast tissues. Overlapping clones 4304443H1 (SEQ ID NO:1), 3040232H1 (SEQ ID NO:2) 3790941H1 (SEQ ID NO:3), 3424294H1 (SEQ ID NO:4) 2741038H1 (SEQ ID NO:5), 4302934H1 (SEQ ID NO:6), 158545H1 (SEQ ID NO:7), respectively, were identified for further study. These represented the minimum number of clones that (along with the full-length sequence of clone 4304443H1 [designated as 4304443inh (SEQ ID NO:8)] were needed to form the contig

and from which the consensus sequence provided herein (SEQ ID NO:9) was derived.

B. Generation of a Consensus Sequence. The nucleotide sequences of clones 4304443H1 (SEQ ID NO:1), 3040232H1 (SEQ ID NO:2) 3790941H1 (SEQ ID NO:3), 3424294H1 (SEQ ID NO:4) 2741038H1 (SEQ ID NO:5), 4302934H1 (SEQ ID NO:6), 158545H1 (SEQ ID NO:7) and the full-length sequence of clone 4304443H1 [designated as 4304443inh (SEQ ID NO:8)] were entered in the Sequencher™ Program (available from Gene Codes Corporation, Ann Arbor, MI) in order to generate a nucleotide alignment (contig map) and then generate their consensus sequence (SEQ ID NO:9). Figures 1A-1E show the nucleotide sequence alignment of these clones and their resultant nucleotide consensus sequence (SEQ ID NO:9). Figure 2 presents the contig map depicting the clones 4304443H1 (SEQ ID NO:1), 3040232H1 (SEQ ID NO:2) 3790941H1 (SEQ ID NO:3), 3424294H1 (SEQ ID NO:4) 2741038H1 (SEQ ID NO:5), 4302934H1 (SEQ ID NO:6), 158545H1 (SEQ ID NO:7) and the full-length sequence of clone 4304443H1 [designated as 4304443inh (SEQ ID NO:8)] that form overlapping regions of the BS322 gene and the resultant consensus nucleotide sequence (SEQ ID NO:9) of these clones in a graphic display. Following this, a three-frame translation was performed on the consensus sequence (SEQ ID NO:9). The third forward frame was found to have an open reading frame encoding a 398-residue amino acid sequence that is presented as SEQ ID NO:24. The open reading frame corresponds to nucleotides 57-1250 of SEQ ID NO:9. A second coding region was found in the second forward reading frame and overlaps the first. This open reading frame (corresponding to nucleotides 1171-2122 of SEQ ID NO:9) encodes a 317-residue amino acid sequence which is presented as SEQ ID NO:25. It is known that rare error in translation, termed translational frameshifting, occur, that allow the ribosome to translate two partially overlapping reading frames as a single polypeptide. I.P. Ivanov et al. RNA 4(10):1230-1238 (1998); and P.J. Farabaugh Annu Rev Genet 30:507-528 (1996). Thus, it is within the scope of this invention that these two partially overlapping reading frames may be translated as such a single polypeptide.

Many other detection formats exist which can be used and/or modified by those skilled in the art to detect the presence of amplified or non-amplified BS322-derived nucleic acid sequences including, but not limited to, ligase chain reaction (LCR, Abbott Laboratories, Abbott Park, IL); Q-beta replicase (Gene-Trak™, Naperville, IL), branched chain reaction (Chiron, Emeryville, CA) and strand displacement assays (Becton Dickinson, Research Triangle Park, NC).

Example 10: Synthetic Peptide Production

Synthetic peptides, SEQ ID NO:26, SEQ ID NO:27 and SEQ ID NO:28, were modeled and prepared based upon the predicted amino acid sequence of the BS322 polypeptide consensus sequence (see Example 1). In particular a number of BS322 peptides derived from SEQ ID NO:24 and SEQ ID NO:25 were prepared, including the peptides of SEQ ID NO:26 and SEQ ID NO:28. All peptides were synthesized on a Symphony Peptide Synthesizer (available from Rainin Instrument Co., Emeryville, CA) or similar instrument, using FMOC chemistry, standard cycles and in-situ HBTU activation. Cleavage and deprotection conditions are as follows: a volume of 2.5 ml of cleavage reagent (77.5% v/v triflouroacetic acid, 15% v/v ethanedithiol, 2.5% v/v water, 5% v/v thioanisole, 1-2% w/v phenol) were added to the resin, and agitated at room temperature for 2-4 hours. The filtrate was then removed and the peptide was precipitated from the cleavage reagent with cold diethyl ether. Each peptide was filtered, purified, via reverse-phase preparative HPLC using a water/acetonitrile/0.1% TFA gradient, and lyophilized. The product was confirmed by mass spectrometry (see Example 12).

Disulfide bond formation is accomplished using auto-oxidation conditions, as follows: the peptide is dissolved in a minimum amount of DMSO (approximately 10 ml) before adding buffer (0.1 M Tris-HCl), pH 6.2) to a concentration of 0.3 – 0.8 mg/ml. The reaction is monitored by HPLC until complete formation of the disulfide bond, followed by reverse-phase preparative HPLC using a water/acetonitrile/0.1% TFA gradient and lyophilization. The product then is confirmed by mass spectrometry (see Example 12).